

REMARKS/ARGUMENTS

1. The amendments made in this reply to the Office Communication of September 20, 2006 have the following objectives:
 - (a) Revision of Claim 1 in response to **Item 7 of the Office Communication**, which objects to the limitation “the application” in step (e), line 2, because there is insufficient antecedent basis for this limitation in the claim;
 - (b) Revision of Claim 1 in response to **Item 8 of the Office Communication**, which objects to the limitation “of any of the word- [weight] setting methods...” in step (e), lines 3-4, because the antecedent basis for the limitation is not clear, considering that said methods have not been presented before step (e) in the claim;
 - (c) Revision of Claim 1 in response to **Item 8 of the Office Communication**, which also objects that it is not clear which step(s) (a)-(e) is implemented in “the computer program Rainbow”, requiring clarification of the metes and bounds.

2. In **Item 7 of the Office Communication** of September 20, 2005, the Examiner objects to the limitation “the application” in step (e), line 2 because there is insufficient antecedent basis for this limitation in the claim. In response, Applicant has removed from step (e) the text that refers to “the application”. In the instant revision, Claim 1, step (e), now begins by claiming “means for running a publicly available computer program known as Rainbow, the running of which is used as a means for annotating ... clusters”. The means for running the program Rainbow is disclosed in the application, for example, on page 37, line 5; on page 38, line 12-13, and more generally on page 18, line 20. Thus, in the preferred embodiment, computer program code of the form

system(“rainbow ...”);

is used to run the computer program Rainbow. As disclosed on page 18, 21ff, in an alternate embodiment, the system could also run Rainbow through use of other programming techniques, such as the use of shell scripts. In the instant revision of Claim 1, step (e), the running of Rainbow as a means for annotating clusters is now made explicit through the following limitations:

said annotating being performed in the following two stages:

- first, instruct the computer program Rainbow to take as extrinsic input-data each and every compendium of text that was constructed in (d) for each individual subset of genes known in (b) as a cluster, then process those input-data to produce as output-data a statistical model of the text in all those compendia; and
- second, instruct the computer program Rainbow to take as its input-data the aforementioned statistical model of the text, and to process that data to produce as output-data a list of words and phrases for each subset of genes known in (b) as a cluster, along with word-weights that Rainbow calculates for each word or phrase in the list, the magnitude of which indicates the weight that the system attaches to the corresponding word or phrase as a characterization of the subset of genes known in (b) as a cluster, said word-weights being calculated by default through Rainbow's implementation of the Naive Bayes algorithm, or optionally through Rainbow's implementation of other word weight-setting algorithms.

The first stage is disclosed on page 37, line 3-8 of the application as follows:

The Text Modeling Module (126) then performs the following system function (written in the C programming language):

system("rainbow -d .\\text_model --index .\\clusters_text*");
to instruct the computer program 'rainbow' to use the text found in subdirectories of the directory \clusters_text, make a statistical model of the text, then store the results in a directory called \text_model [McCALLUM, supra].

The second stage is disclosed, for example, on page 38, lines 10-18 of the application as follows:

The following system function, written in the C programming language, is used by the Keyword Identification Module (128) to generate the word lists:

system("rainbow -d .\\text_model --print-word-weights=cluster_name>word-weights_cluster_name.txt");
where cluster_name is the name of a cluster (0,1, 2,..., or Cmax) and where word-weights_cluster_name.txt is the name of a text file that is to contain the words and corresponding weights for that particular cluster. Word lists for each of the clusters are generated in succession by performing a system command of this form, but with different cluster names (i.e., cluster numbers).

The default algorithm that is used to calculate these word weights (Naive Bayes), as well as the option of selecting alternate algorithms, are disclosed on page 41, line 23 through page 42, line 12 of the application. See also Amendment 27 in the Applicant's response of October 20, 2003 to the Office Action with mailing date of April 22, 2003, which moved the location of this text. The generation of words listed in ranked order according to the magnitude of their corresponding log-odds ratio scores and of their mutual information are also disclosed on page 39, lines 3-15.

3. In Item 8 of the Office communication of September 20, 2005, the Examiner objects to the limitation “of any of the word- [weight] setting methods...” in step (e), lines 3-4, because the antecedent basis for the limitation is not clear, considering that said methods have not been presented before step (e) in the claim. In response to this objection, the instant revision of Claim 1, step (e) makes clear that running of the publicly-available computer program Rainbow results in calculation of the word-weights (considered simply as an input/output procedure, irrespective of the details of how Rainbow actually calculates the word-weights). It also makes clear that by default, Rainbow uses the Naive Bayes algorithm to calculate those word-weights and that optionally, Rainbow may be instructed to use other algorithms for that purpose. Such algorithms are disclosed on page 41, line 23 through page 42, line 12 of the application. See also Amendment 27 in the Applicant’s response of October 20, 2003 to the Office Action with mailing date of April 22, 2003, which moved the location of this text. As disclosed there and in the “rainbow --help” documentation under “--method=METHOD”, optional algorithms include support vector machines (svm), term frequency- inverse document frequency (tfidf), probabilistic indexing (prind), maximum entropy (maxent), k-nearest neighbors (knn), EM algorithm (em), Dirichileit kernel (dirk), and Active Learning (active).

Means for generating words lists, ranked in order according to the magnitude of the words’ corresponding log-odds ratio scores and of their mutual information are also disclosed on page 39, lines 3-15. As disclosed there, when Rainbow calculates the log-odds ratio scores or mutual information scores for all the words in the text compendia, it then proceeds to sort and prune the words (e.g., to the top 20) according to their rank, as evaluated in terms of the magnitude of their corresponding scores. In contrast, when Rainbow is instructed to produce the word lists with corresponding weights calculated according to the Naive Bayes or the other previously-mentioned classification algorithms, in general, it does not sort or prune those word lists, but instead produces a voluminous list of words from the text compendium, which are not in general ordered according to the magnitude of the corresponding weights.

Therefore, the Claims includes a final step (f) to claim the sorting and pruning, which produces the desired ordering of words and their corresponding weights, irrespective of which of the algorithms was invoked through Rainbow to produce the initial word list. Such a step was disclosed on page 39, line 21 through page 40, line 2 of the application, where it is stated that “by

default, only words with positive weights are read into the Key Words or Phrases section (158), sorted in descending order according to the weight values. As an option, words with negative weights may be included as well". Tables 1 and 2 give examples of the word lists and corresponding weights, produced through operation of the Data Output Module (136), which displays the key words or phrases (application page 44, lines 19-21). The standard sorting algorithm that is used by the system was disclosed on page 24, line 21 as being published in Press et al (1992), which consists of C-language code programming recipes. The disclosure of pruning is inherent in the fact that by default, only words with positive word-weights are read into the Key Words of Phrases section (i.e., pruning words having corresponding weights less than zero) and in the fact that Tables 1 and 2 show the top 25 keywords (i.e., pruning words having corresponding weights less than that of the word at the 25th rank). Note that the pruning described in the instant Claim 1, step (f) is concerned only with ignoring words or phrases having corresponding weight values less than a cutoff value, which is different than the pruning that is disclosed on page 38, lines 1-3, which is to ignore words that occur less than some specified number of times in the text files, or the pruning of words that occur in a stoplist of common words (application, page 37, lines 12-13).

Applicant believes that under the doctrine of inherency, Rainbow's implementation of the above-mentioned algorithms are a part of the original disclosure because they were identified in the original disclosure (application page 41, line 23 through page 42, line 12; and page 39, lines 3-15). The doctrine of inherency states that if an application discloses an object [e.g., source code and instructional documentation for the "rainbow" computer program] that inherently includes a function, property, or advantage [e.g., implementation of the above-named algorithms], then that function, property or advantage is necessarily disclosed by that application even though the application says nothing directly about the inherent function, property or advantage. (Technicon Instruments Corp. v. Cole Instrument Inc., 255 F. Supp. 630, 640-41, 7th Cir. 1966).

4. In Item 8 of the Office communication of September 20, 2005, the Examiner objects that it is not clear which step(s) (a)-(e) is implemented in "the computer program Rainbow", requiring clarification of the metes and bounds. In response, in the instant revision, Claim 1, step (e), now begins by claiming "means for running a publicly available computer program known as Rainbow,

the running of which is used as a means for annotating ... clusters". It is implicit from this revision of Claim 1 that the computer program Rainbow is not used in steps (a) through (d), otherwise one would have claimed the means for running Rainbow in a step earlier than (e). It is also implicit from the revision of Claim 1, step (e) that Rainbow was not used in an earlier step, because it is stated in step (e) that the compendia of text constructed in step (d) are used as extrinsic input-data for processing by Rainbow. If the compendia of text constructed in (d) had been generated (directly or indirectly) through use of the Rainbow computer program, then they would not have been extrinsic to Rainbow. The fact that these input-data are extrinsic to Rainbow is also made clear in the disclosure (application, pages 21 through 36), because there is no mention of Rainbow's use in generating the text compendia that are used as its input-data.

5. The Office communication of September 20, 2005 does not raise any objections pertaining to the disclosure, as currently amended, apart from the Claims. Therefore, as a courtesy to the Examiner, Applicant encloses with this response a clean copy of the disclosure (without the claims), incorporating all of its amendments up to this date.

6. In Item 9 of the Office Communication of September 20, 2005, the Examiner indicates that Claim 1 would be allowable if rewritten or amended to overcome the objections indicated in Items 7 and 8 of the Office Communication. Applicant believes that the instant amendments to Claim 1 are responsive to those objections. However, if the Examiner finds patentable subject matter disclosed in this application, but still feels that Applicant's present claims are not entirely suitable, Applicant requests pursuant to MPEP 707.07(j) that the Examiner draft one or more suitable claims for the applicant.

Respectfully submitted,

David R. Rigney

David R. Rigney, Inventor

GENETWORKS Inc.
P.O. Box 33296
Austin TX 78764-0296